PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : H04L 12/56, H04Q 11/04

A1

(11) International Publication Number: WO 99/66677

(43) International Publication Date: 23 December 1999 (23.12.99)

(21) International Application Number: PCT/EP99/04376

(22) International Filing Date: 15 June 1999 (15.06.99)

(30) Priority Data:
09/098,228 16 June 1998 (16.06.98) US

(71) Applicant: ALCATEL [FR/FR]; 54, rue La Boétie, F-75008 Paris (FR).

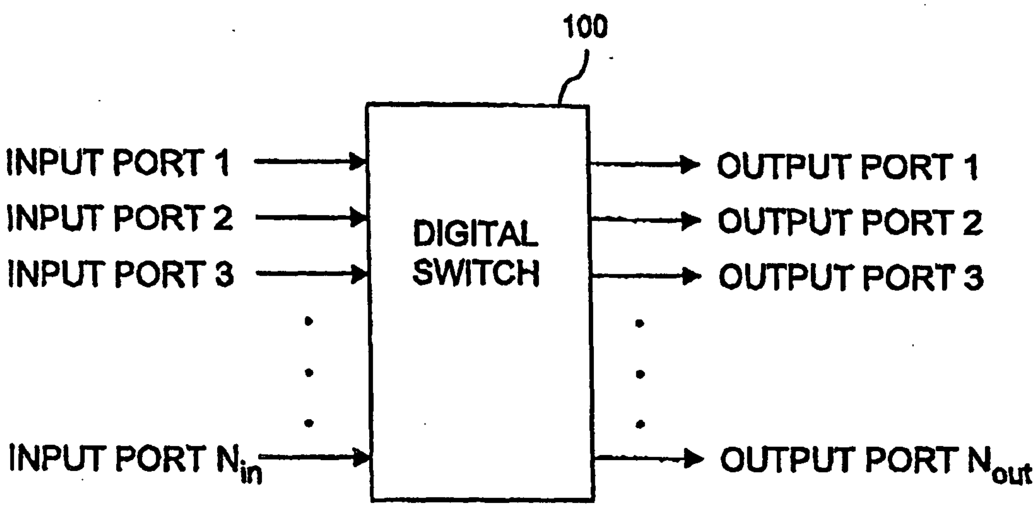
(72) Inventors: HOOGENBOOM, Chris, L.; 2230 Peninsula Road, Channel Islands Harbor, CA 93035 (US). HILL, Rex, A.; 9134C Regents Road, La Jolla, CA 92037 (US).

(74) Agents: RAUSCH, Gabriele et al.; Alcatel, Intellectual Property Dept. Stg., Postfach 300929, D-70449 Stuttgart (DE).

(81) Designated States: CN, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: DIGITAL TRAFFIC SWITCH WITH CREDIT-BASED BUFFER CONTROL



(57) Abstract

A credit-based buffer control scheme for an input-buffered/output-controlled switch allocates turns to release protocol data units deterministically among transmit queues storing groups of protocol data units entitled to different qualities of service. Receive queues may be constructed dynamically within a shared output buffer to accommodate non-uniform traffic patterns. In a LAN or WAN packet switch implementation, transmit queue data release rates over a period of sustained operation are made independent of the length of protocol data units presented for release by employing a "burst"-based release strategy.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

DIGITAL TRAFFIC SWITCH WITH CREDIT-BASED BUFFER CONTROL

BACKGROUND OF THE INVENTION

The present invention relates to digital traffic switches, and more particularly to input-buffered/output-controlled digital traffic switches.

A digital traffic switch takes protocol data units from many sources, and routes them to many destinations. Such a switch may be hardware-based, or hardware- and software-based. The architecture can be generalized as shown in Figure 1. Protocol data units arrive on input ports 1 to N_{in} and are switched through switch 100 to various ones of output ports 1 to N_{out}.

Sometimes, collisions take place between protocol data units. That is, protocol data units come in on two or more input ports that are destined for the same output port. There are two scenarios which describe the condition when two or more protocol data units destined for the same output port collide: (1) momentary contention or collision; and (2) sustained contention or congestion. To minimize collisions, buffering is used. Buffers temporarily store protocol data units while other protocol data units are consuming the resources necessary for delivery of the buffered protocol data unit to the output port.

Buffering may give rise to a another traffic flow problem known as blocking. Blocking occurs when delivery of a protocol data unit is delayed indefinitely because the delivery of other protocol data units is consuming the resources necessary for the delivery of the blocked protocol data unit to the output port. This is referred to as "head of line" blocking. Blocking is undesirable, since it can delay the delivery

of the blocked protocol data units and cause them to be overwritten while awaiting release in the buffers, a condition known as "dropping".

Blocking may take several different forms. In asynchronous transfer mode (ATM) switches, blocking may prevent all protocol data units having a relatively low priority from being delivered to the output side, a condition known as "priority blocking". In the case of LAN and WAN packet switches, blocking may prevent all protocol data units having a particular set of packet identifiers from being delivered to the output side, a condition called "flow blocking". And in the case of ATM, LAN or WAN, blocking may prevent all protocol data units arriving on a particular input port from being delivered to the output side, a condition called "port blocking".

To prevent blocking, a switch must implement a strategy for allocating the limited bandwidth available for transmitting protocol data units from the input ports to the output ports. One important aspect of bandwidth allocation is buffer control. Buffer control times the release of protocol data units from buffers for delivery to other buffers on the path between input ports and output ports.

One buffer control strategy buffers protocol data units at the input side and throttles the input buffers based on the availability of output port bandwidth. This is known as input-buffered/output-controlled switching. Protocol data units received at the input side and destined for the output side are buffered at the input side. Traffic load on the output ports is monitored and the protocol data units are released to the output side when the load is sufficiently light. Releasing protocol data units buffered at inputs based on the load status of outputs has advantages. First, input buffering means that protocol data units are dropped before being transmitted

across the switch fabric whenever dropping is required, reducing the traffic load on the switch fabric during congested periods. Second, output-controlled release means that the switch may be designed around output port bandwidth limitations, improving scalability. Third, output-controlled release may be used to normalize traffic patterns, i.e., traffic flow on the switch may be accelerated during periods of low traffic and decelerated during periods of high traffic.

Input-buffered/output-controlled ATM switches are known. In such switches, protocol data units are fixed-length "cells" whose release across the switch fabric is usually triggered by a handshaking procedure in which "requests" and "grants" are exchanged on a "per cell" basis. More particularly, the input side buffers inbound cells and transmits to the output side a request to release each input-buffered cell. The request typically specifies type information for the cell, e.g., source input, cell priority, destination output, etc. The output side eventually issues a grant to release the cell across the switching fabric to the destination output port. The ordering of grants has typically been made by running an algorithm which arbitrates among the pending requests based on their specified type whenever bandwidth becomes available. Such switches have eliminated port blocking by running algorithms that issue grants round-robin among input ports, see, e.g., Khacherian, U.S. Patent No. 5,768,257.

Priority blocking, however, has been allowed to persist. Grants to release relatively high priority cells have been issued before grants to release relatively low priority cells. Therefore, whenever the switch is flooded with relatively high priority traffic, no relatively low priority traffic has been delivered.

Clearly, a round-robin allocation of "turns" (e.g., grants) to release among different priorities would not make sense. Prioritization is inherently non-egalitarian: It is supposed to create a "quality of service" hierarchy which provides certain protocol data units more favorable treatment than others. Still, the "all or nothing" approach to allocating turns to release used in known input-buffered/output-controlled ATM switches, which has allowed priority blocking to persist, is not wholly satisfactory. A more sophisticated scheme for allocating turns to release among different priorities is warranted. Such a scheme is also warranted in input-buffered/output-controlled LAN and WAN packet switches to prevent the analogous problem of flow blocking. Stating the problem generally, a better scheme is needed in input-buffered/output-controlled switches for allocating turns among protocol data units entitled to different qualities of service.

SUMMARY OF THE INVENTION

In its most basic feature, the present invention provides, in an input-buffered/output-controlled switch, a credit-based scheme that allocates turns to release deterministically among groups of protocol data units entitled to different qualities of service.

In one aspect, the input-buffered/output-controlled switch has one or more input units each associated with one or more transmit queues and one or more output ports each associated with one or more receive queues. The input units transmit data to the output ports over a shared switching fabric. Inbound protocol data units destined for different output ports or entitled to different qualities of service are assigned to different transmit queues. The traffic load on receive queues

is monitored. Whenever the traffic load on a receive queue is sufficiently light, the receive queue declares a "pay day" by sending a status message to the input units. In response to the "pay day" declaration, the input units distribute "paychecks" to transmit queues which transmit protocol data units to the declaring receive queue. "Paychecks" distributed to different transmit queues contain different amounts of credit depending on the quality of service entitlement of the protocol data units assigned to the transmit queue. Transmit queues are polled in a predetermined order for releasing protocol data units to the receive queue. If a transmit queue has a sufficient of amount of credit when polled, the transmit queue releases a protocol data unit to the receive queue and the amount of credit accumulated by the transmit queue is reduced in relation to the length of the released protocol data unit. If a transmit queue does not have a sufficient amount of credit when polled, the transmit queue does not release a protocol data unit to the receive queue and retains its accumulated credits. The amount of credit distributed to transmit queues in "paychecks" may be arranged such that transmit queues storing protocol data units entitled to a relatively high quality of service are distributed a relatively large amount of credit while transmit queues storing protocol data units entitled to a relatively low quality of service are distributed a threshold amount of credit which is relatively small. By allocating credit in this way, a quality of service hierarchy is preserved while blocking is prevented.

In another aspect, for implementation in a LAN or WAN packet switch, a standard "burst" count is selected for all transmit queues and assigned to the transmit queue whose "turn" it is to release within the predetermined order. The

transmit queue releases protocol data units until the "burst" count has been drawn-down to a level where the length of the next packet for release exceeds the remaining "burst" count. The residual portion of the assigned "burst" count, if any, is stored and assigned as extra "burst" on the transmit queue's next "turn" to release. Transmit queue data release rates over a period of sustained operation are thereby made independent of the length of protocol data units presented for release.

In yet another aspect, the receive queues are constructed dynamically within a shared output buffer to accommodate non-uniform traffic patterns.

These and other objects of the present invention may be better understood by reference to the following detailed description, taken in conjunction with the accompanying drawings which are briefly described below. Of course, the actual scope of the invention is defined by the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a generalized block diagram of a digital traffic switch;

Figure 2 is a block diagram of a digital traffic switch in which the present invention may be used;

Figure 3 is a block diagram showing in greater detail the control flow between the output flow controller and the input buffer controller, as well as the data flow between the input buffer and the output buffer;

Figure 4 is a flow diagram of a load control algorithm used in a preferred embodiment to declare "pay days"; and

Figure 5 is a flow diagram of a release control algorithm used in a preferred embodiment to order the release protocol data units from the transmit queues to the receive queues.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring to Figure 2, the general architecture of a digital traffic switch incorporating the present invention is shown. Multiple input units $210_{(1-N)}$ and multiple output units $220_{(1-N)}$ are coupled via a switch fabric 230. Each input unit may have one or more associated input ports and each output unit may have one or more associated output ports such that every input port may communicate with every output port. At any given instant in time, a subset (or all) of the input units $210_{(1-N)}$ receive digital traffic destined for a subset (or all) of the output units $220_{(1-N)}$. Digital data traffic may therefore be imagined as flowing from left to right, from input units $210_{(1-N)}$ to output units $220_{(1-N)}$. Data traffic is transmitted in discrete protocol data units. Depending on the protocol operative on the switch, protocol data units may be either variable-length packets or fixed-length cells.

Figure 3 isolates the connections across switching fabric 330 between a representative input unit 310 and a representative output unit 340. Within input unit 310, there is an input buffer 312 which has different transmit queues for storing inbound protocol data units destined for different output ports or entitled to different qualities of service. For a switch supporting L different qualities of service and having U output units each supporting P output ports, the input buffer will have U x P x L transmit queues (residing physically within the input buffer). There is also an input buffer controller 314 which assigns inbound protocol data units to different

transmit queues by output port and quality of service entitlement, monitors the available credits for different transmit queues and causes protocol data units to be released from input buffer 312 to output buffer 332. Quality of service may be determined for an inbound protocol data unit by reference to a predetermined set of switching characteristics. In the case of a LAN or WAN packet switch, the determinative quality of service characteristics may include, without limitation, control and routing information found in Layer 2 (media access), Layer 3 (network) and Layer 4 (transport) headers of the protocol data unit, alone or in combination. In the case of an ATM cell switch, determinative quality of service characteristics may include, without limitation, input port and cell priority. Within input buffer controller 314, available credit stores 316 retain the current credit values for different transmit queues. For an input unit supporting U x P x L transmit queues, there will be U x P x L available credit stores each retaining a credit value for a different one of the U x P x L transmit queues.

Within switching fabric 330, there is an output buffer 332 which has different receive queues for storing protocol data units destined for different output ports. For a switch having U output units each supporting P output ports, the output buffer will have U x P logical receive queues (residing physically within the output buffer). Within switch fabric 330, there is also an output flow controller 334 which monitors the traffic load on the different receive queues and causes "pay day" declarations to be issued accordingly. Output flow controller 334 has credit regulation stores 336 for monitoring the current traffic load on the different receive queues. For a switching fabric supporting U x P receive queues, there will be U x P credit

regulation stores each retaining a load value for a different one of the receive queues.

Within output unit 340, there are different output FIFOs 342 for storing protocol data units destined for different output ports.

In a preferred embodiment, the resources of output buffer 332 are allocated dynamically with the expedient of a free buffer pointer list in output flow controller 334. Pointers to available address spaces within output buffer 332 are pulled from the front of the free list when a protocol data unit is written to output buffer 332 and are returned to the back of the free list when a protocol data unit is read from output buffer 332. Address spaces holding protocol data units (or portions thereof) destined for the same output port are linked to form a logical receive queues. By employing a shared output buffer in which receive queues are constructed dynamically, non-uniform traffic patterns may be advantageously accommodated by allowing output ports to utilize more than their proportionate share of output buffer 332. Of course, a dedicated output buffering environment wherein each output unit has its own output buffer and output flow controller is also possible. In that event, each output buffer would have P receive queues and P credit regulation stores corresponding to its P output ports. Such a dedicated output buffering environment may have certain scalability advantages relative to a shared output buffering scheme.

Input unit 310 and switching fabric 330 communicate on status line 318 and data bus 320. Status line 318 is used to transmit "pay day" declarations from output flow controller 334 to input buffer controller 314. Each "pay day"

declaration includes information sufficient to identify the receive queue to which the "pay day" declaration pertains. A "pay day" declaration is transmitted when the traffic load on a receive queue is found to be below a predetermined threshold. In a preferred embodiment, traffic load is measured by the amount of data in the receive queue awaiting delivery to the receive queue's corresponding output port. "Pay day" declarations may be transmitted in various ways. In a preferred embodiment, status line 318 is a multi-bit bus for transmitting up to one multi-bit "pay day" declaration on each clock cycle. In an alternative embodiment, each receive queue may be assigned a dedicated one-bit line for conveying "pay day" declarations by pulsing or failing to pulse the line high. In yet another alternative embodiment, receive queues may be assigned to one of a plurality of shared one-bit lines for conveying "pay day" declarations and the receive queues sharing a line may be assigned different time slots in a repetitive timing cycle for conveying "pay day" declarations by pulsing or failing to pulse the line on their assigned time slots. Of course, the number of receive queues may determine the relative suitability of the various available techniques for conveying "pay day" declarations. Data bus 320 is used to transmit protocol data units from transmit queues to receive queues.

Output flow controller 334 updates appropriate load values when protocol data units are transmitted from input buffer 312 to output buffer 332 and from output buffer 332 to output FIFOs 342. More particularly, when a protocol data unit is received in a receive queue within output buffer 332, output flow controller 334 increases the load value retained for the corresponding receive queue in relation to the length of the received protocol data unit. When a protocol data unit

is transmitted from a receive queue within output buffer 332, output flow controller 334 decreases the load value retained for the corresponding receive queue in relation to the length of the transmitted protocol data unit. In this way, the output flow controller 334 dynamically maintains an accurate view of the traffic load on the various receive queues within output buffer 332. The relationship between protocol data unit length and the load value may be a direct proportion, such that each N-bit data block written to the receive queue may increase the load value by one, and each N-bit data block read from the receive queue may reduce the load value by one.

In a similar vein, input buffer controller 314 updates the appropriate credit values when "pay day" declarations are received from output flow controller 334 and when protocol data units are transmitted from input buffer 312 to output buffer 332. More particularly, when a "pay day" declaration is received, input buffer controller 314 increases the credit values for the transmit queues which transmit protocol data units to the receive queue that transmitted the "pay day" declaration. Credit values are increased in relation to the predetermined quality of service entitlement of the protocol data units assigned to the transmit queue. The amount of credit distributed is arranged such that transmit queues assigned protocol data units entitled to a relatively high quality of service are distributed a relatively large number of credits while transmit queues assigned protocol data units entitled to a relatively low quality of service are distributed a threshold number of credits which is relatively small. When a protocol data unit is transmitted from a transmit queue to a receive queue within output buffer 332, input buffer controller 314 decreases the

credit value for the corresponding transmit queue in relation to the length of the transmitted protocol data unit. In this way, the input buffer controller 314 keeps an accurate view of the credits held by the various transmit queues within input buffer 312. The relationship between protocol data unit length and the credit value may be a direct proportion, such that each N-bit data block read from the transmit queue may reduce the credit value by one.

The general buffer control strategy operative in the switch will now be described by reference to the interaction of representative input unit 310 and switch fabric 330. Output flow controller 334 polls receive queues round-robin. "Pay day" declarations are transmitted to input buffer controller 314 when the load on the current receive queue is sufficiently light. The load value for a receive queue is retained in the receive queue's one of credit regulation stores 336. In a preferred embodiment, the load on a receive queue is deemed sufficiently light when the average load value over a predetermined number of clock cycles is below a predetermined threshold. In response to "pay day" declarations, input buffer controller 314 distributes "paychecks" to appropriate transmit queues. Each transmit queue which directs protocol data units to the declaring receive queue is distributed a "paycheck". However, not all "paychecks" have the same amount of credit. The amount of credit in a "paycheck" correlates with the quality of service entitlement of protocol data units assigned to the transmit queue. Input buffer controller 314 deposits "paychecks" in the available credit stores 316 by revising the credit values retained for the appropriate transmit queues upward by the amount of the credit indicated in the "paychecks". Meanwhile, input buffer

controller 310 determines the output port and the quality of service entitlement for inbound protocol data units and inbound protocol data units are assigned to transmit queues accordingly. Input buffer controller 310 polls transmit queues round-robin. A protocol data unit is released to the appropriate receive queues from current transmit queue, provided the current transmit queue has a sufficiently large amount of credit. In response to receipt of a protocol data unit, output buffer controller 322 revises the load values retained for the appropriate receive queues upward by the length of the received protocol data unit. Meanwhile, protocol data units are transmitted from receive queues to output FIFOs 342 en route to output ports in a predetermined order, as output port bandwidth becomes available.

Various enhancements of this basic buffer control strategy are possible. Nevertheless, at a fundamental level, this basic buffer strategy is believed to represent a significant advance over the prior art in that "turns" to release protocol data units among transmit queues may be allocated in a wholly deterministic manner. Thus, in a preferred embodiment, credits may be distributed in a manner which ensures that differences in quality of service are amply respected, but prevents blocking. This advantage may be achieved by distributing a higher number of credits to transfer queues assigned protocol data units entitled to relatively high quality of service than to transfer queues assigned protocol data units entitled to relatively low quality of service, while assigning a threshold number of credits to transfer queues assigned protocol data units entitled to the lowest quality of service.

One preferred enhancement to this basic buffer control strategy maximum imposes maximum forwarding rate limits by requiring a predetermined number of

clock cycles to expire between "pay check" distributions and setting judicious caps on the number of credits distributed per "pay check" to any one transfer queue.

Another preferred enhancement to the basic buffer control strategy improves fairness in a LAN or WAN packet switch through "burst"-implemented release. A standard "burst" count is selected for all transmit queues and the "burst" count is assigned to the transmit queue whose "turn" it is to release in the round-robin order. The transmit queue releases protocol data units, in this case packets, until the "burst" count has been drawn-down to a level where the length of the next packet for release exceeds the remaining "burst" count. The residual portion of the assigned "burst" count, if any, is stored in a residual "burst" store retained for the transmit queue and is reassigned as extra "burst" on the transmit queue's next "turn" to release. On subsequent "turns", the transmit queue releases packets until the "burst" count plus any residual "burst" from the previous "turn" has been drawn-down. Preferably, the selected "burst" count is large enough to guarantee that at least one packet is released on every "turn". By implementing the foregoing strategy, transmit queue data release rates over a period of sustained operation are made independent of the length of packets presented for release.

Referring now to Figure 4, the load control algorithm implemented by the output flow controller to declare "pay days" is shown. The current receive queue in a round-robin order is selected (410) and the current traffic load on the receive queue is reviewed (420). If the current traffic load is sufficiently light, a "pay day" is declared (430). Whether or not the current traffic load is sufficiently light, a check is made to determine if the current receive queue is the last receive queue (440). If it

is, the algorithm is exited (450). If it is not, the next receive queue in the round-robin order is selected (410) and the Step 420 is repeated.

Referring to Figure 5, the release control algorithm implemented by the input buffer controller to release protocol data units is shown. The current transmit queue in a round-robin order is selected (510) and the current credit value for the transmit queue is reviewed (520). If the current credit value is sufficiently high, a protocol data unit is released (530) and the credit value is reduced in relation to the length of the released protocol data unit (540). Whether or not the current credit value is sufficiently high, a check is made to determine if the current transmit queue is the last transmit queue (550). If it is, the algorithm is exited (450). If it is not, the next transmit queue in the round-robin order is selected (510) and the Step 520 is repeated.

The algorithms may be implemented in software using a general purpose microprocessor. More preferably, the algorithms may be implemented in hardware as part of an application specific integrated circuit (ASIC). In the latter instance, the flow diagrams of Figures 4 and 5 describe not the execution of program instructions but the logical operation of the ASIC.

Polling orders other than round-robin polling among transfer queues are possible. For example, transmit queues may be polled according to the relative priority of their assigned protocol data units, such that relatively high priority protocol data units are released before relatively low priority protocol data units until the transmit queues assigned relatively high priority protocol data units no longer have sufficient credit. Priority blocking may be prevented in such an

arrangement by imposing maximum forwarding rate limits on transfer queues assigned relatively high priority protocol data units, in a manner such as that heretofore described.

Also, more sophisticated schemes for declaring "pay days" may be implemented. For instance, a plurality of priority-specific thresholds may be assigned to each receive queue, with higher priorities assigned higher thresholds. When the load value exceeds at least one but not all of the thresholds, a selective "pay day" may be declared resulting in the distribution of "pay checks" only to the transmit queues associated with the receive queue that are assigned protocol data units of a priority whose threshold has not been exceeded.

Therefore, it will be appreciated by those of ordinary skill in the art that the invention can be embodied in other specific forms without departing form the spirit or essential character hereof. The present description is therefore considered in all respects illustrative and not restrictive. The scope of the invention is indicated by the appended claims, and all changes that come within the meaning and range of equivalents thereof are intended to be embraced therein.

Claims

1. A method of controlling buffering of protocol data units, for use within a digital switch having multiple input ports, multiple output ports, and a switch fabric for switching protocol data units received at any of the input ports to any of the output ports, wherein each input port is associated with an input buffer having one or more transmit queues and wherein each output port is associated with an output buffer having one or more receive queues, the method comprising:

(a) receiving protocol data units on input ports, the protocol data units each destined for a particular output port;

(b) assigning the protocol data units to transmit queues, the transmit queues each arranged for storing protocol data units destined for a particular output port;

(c) polling receive queues in a predetermined order, the receive queues each arranged for storing protocol data units destined for a particular output port;

(d) if the load on a polled receive queue is sufficiently light, assigning credit to each transmit queue associated with the same output port as the polled receive queue;

(e) polling the transmit queues in a predetermined order;

(f) if the polled transmit queue has a sufficient of amount of credit, transmitting a protocol data unit from the polled transmit queue to the

receive queue associated with the same output port as the polled transmit queue; and

(g) further transmitting the transmitted protocol data unit from the receive queue to the destination output port.

2. The method according to claim 1, wherein step (f) further comprises reducing the amount of credit assigned to the polled transmit queue in relation to the length of the transmitted protocol data unit.

3. The method according to claim 1, wherein in step (d) the amount of credit assigned varies among transmit queues.

4. The method according to claim 1, wherein in step (b) each transmit queue is arranged for storing protocol data units entitled to a particular quality of service.

5. The method according to claim 4, wherein in step (d) the amount of credit assigned varies among transmit queues in relation to the quality of service to which protocol data units stored in the transmit queue are entitled.

6. The method according to claim 4, wherein in step (d) transmit queues for storing protocol data units entitled to a relatively high quality of service are assigned more credit than transmit queues for storing protocol data units entitled to a relatively low quality of service.

7. The method according to claim 1, wherein in step (d) each transmit queue is assigned a threshold amount of credit.

8. The method according to claim 1, wherein in step (d) the amount of credit assigned is configurable.

9. The method according to claim 1, wherein in step (d) the load is measured by comparing the amount of data awaiting transfer to the output port against a predetermined threshold.

10. The method according to claim 1, wherein in step (c) the receive queues are polled round-robin.

11. The method according to claim 1, wherein in step (e) the transmit queues are polled round-robin.

12. A method of controlling buffering of protocol data units, for use within a digital switch having multiple input ports, multiple output ports, and a switch fabric for switching protocol data units received at any of the input ports to any of the output ports, wherein each input port is associated with an input buffer having one or more transmit queues and wherein each output port is associated with an output buffer having one or more receive queues, the method comprising:

(a) receiving protocol data units on input ports;

(b) assigning the protocol data units to transmit queues;

(c) polling receive queues in a predetermined order, the receive queues each comprising a linked list of entries within a shared output buffer arranged to receive protocol data units from particular transmit queues and to transmit the protocol data units to a particular destination output port;

(d) if the number of entries in a polled receive queue is sufficiently small, assigning credit to each of the particular transmit queues arranged to transmit protocol data units to the polled receive queue;

(e) polling the transmit queues in a predetermined order;

(f) if a polled transmit queue has a sufficient of amount of credit, transmitting a protocol data unit from the polled transmit queue to the receive queue arranged to receive protocol data units from the polled transmit queue; and

(g) further transmitting the transmitted protocol data unit from the receive queue to the destination output port.

13. The method according to claim 12, wherein step (f) further comprises reducing the amount of credit assigned to the polled transmit queue in relation to the length of the transmitted protocol data unit.

14. The method according to claim 12, wherein in step (d) the amount of credit assigned varies among transmit queues.

15. The method according to claim 12, wherein in step (b) each transmit queue is arranged for storing protocol data units entitled to a particular quality of service.

16. The method according to claim 15, wherein in step (d) the amount of credit assigned varies among transmit queues in relation to the quality of service to which protocol data units stored in the transmit queue are entitled.

17. The method according to claim 15, wherein in step (d) transmit queues for storing protocol data units entitled to a relatively high quality of service are assigned more credit than transmit queues for storing protocol data units entitled to a relatively low quality of service.

18. The method according to claim 12, wherein in step (d) each transmit queue is assigned a threshold amount of credit.

19. The method according to claim 12, wherein in step (d) the amount of credit assigned is configurable.

20. The method according to claim 12, wherein in step (c) the receive queues are polled round-robin.

21. The method according to claim 12, wherein in step (e) the transmit queues are polled round-robin.

22. A digital switch of the type having multiple input ports, multiple output ports and a switch fabric for switching protocol data units received at any of the input ports to any of the output ports, comprising:

an input buffer associated with each input port, the input buffer having one or more transmit queues;

an output buffer having one or more receive queues;

means for receiving protocol data units on the input ports, the protocol data units each destined for a particular output port;

means for assigning the protocol data units to transmit queues, the transmit queues each arranged for storing protocol data units destined for a particular output port;

means for polling receive queues in a predetermined order, the receive queues each arranged for storing protocol data units destined for a particular output port;

means for assigning credit to each transmit queue associated with the same output port as a polled receive queue having a sufficiently light load;

means for polling the transmit queues in a predetermined order;

means for transmitting a protocol data unit from a polled transmit queue having a sufficient amount of credit to the receive queue associated with the same output port as the polled transmit queue and for reducing the amount of credit assigned to the polled transmit queue in relation to the length of the transmitted protocol data unit; and

means for further transmitting the transmitted protocol data unit from the receive queue to the destination output port.

Fig. 1

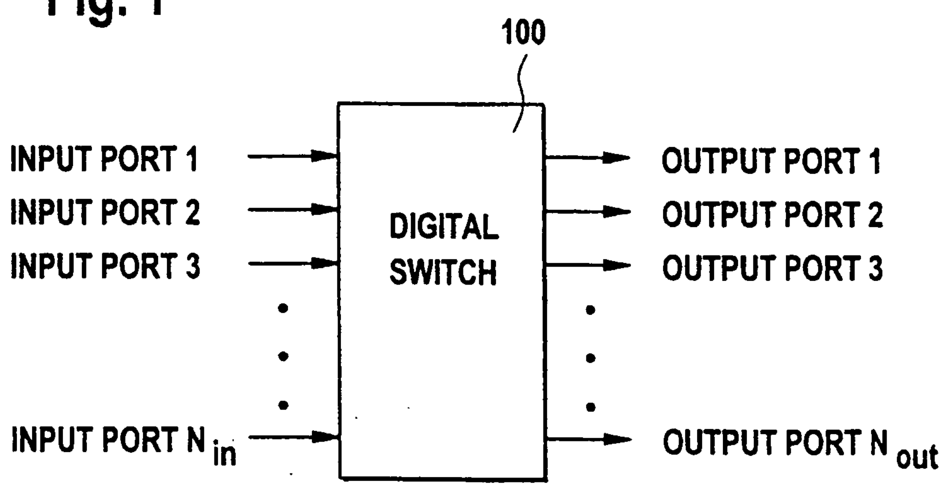

100
INPUT PORT 1
INPUT PORT 2
INPUT PORT 3
INPUT PORT N in
DIGITAL SWITCH
OUTPUT PORT 1
OUTPUT PORT 2
OUTPUT PORT 3
OUTPUT PORT N out

Fig. 2

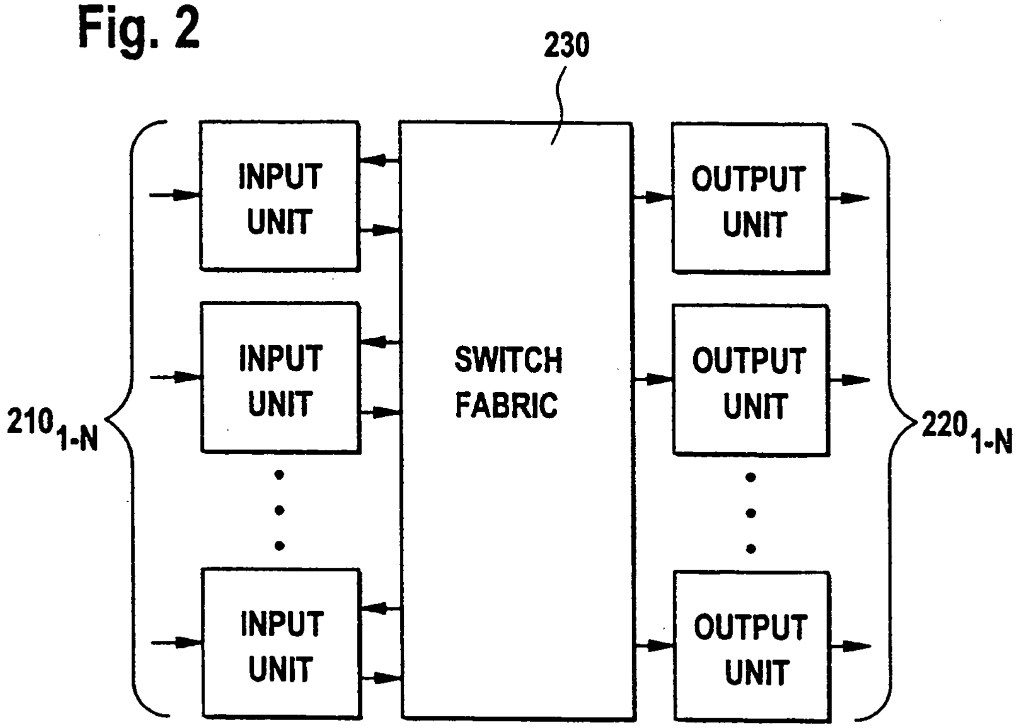

230
INPUT UNIT
INPUT UNIT
INPUT UNIT
SWITCH FABRIC
OUTPUT UNIT
OUTPUT UNIT
OUTPUT UNIT
210 1-N
220 1-N

Fig. 3

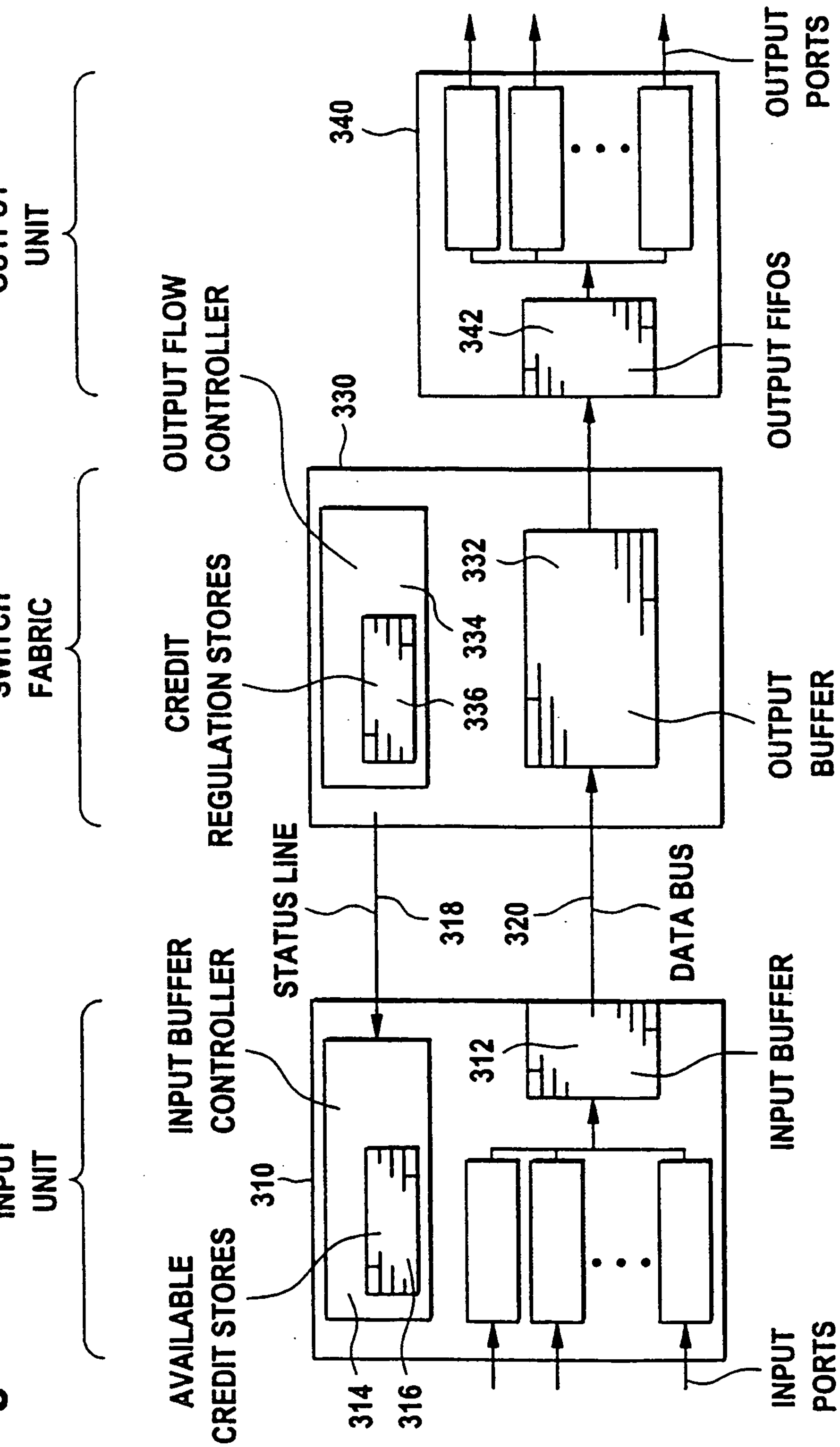
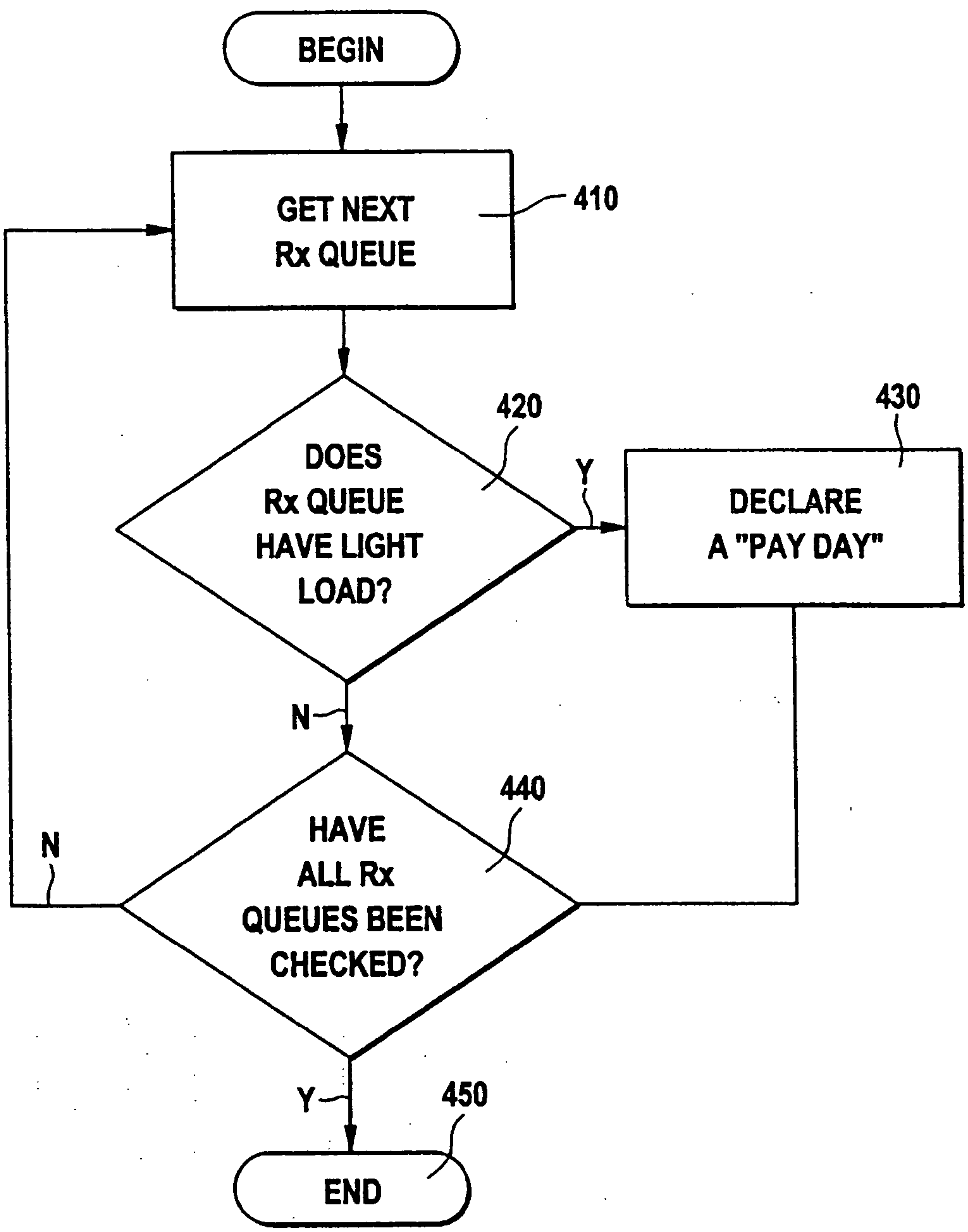

INPUT UNIT
SWITCH FABRIC
OUTPUT UNIT
AVAILABLE CREDIT STORES
INPUT BUFFER CONTROLLER
310
314
316
312
INPUT BUFFER
INPUT PORTS
STATUS LINE
318
320
DATA BUS
CREDIT REGULATION STORES
OUTPUT FLOW CONTROLLER
330
332
334
336
OUTPUT BUFFER
340
342
OUTPUT FIFOS
OUTPUT PORTS

Fig. 4


BEGIN
GET NEXT
Rx QUEUE
410
DOES
Rx QUEUE
HAVE LIGHT
LOAD?
420
Y
430
DECLARE
A "PAY DAY"
N
HAVE
ALL Rx
QUEUES BEEN
CHECKED?
440
N
Y
450
END

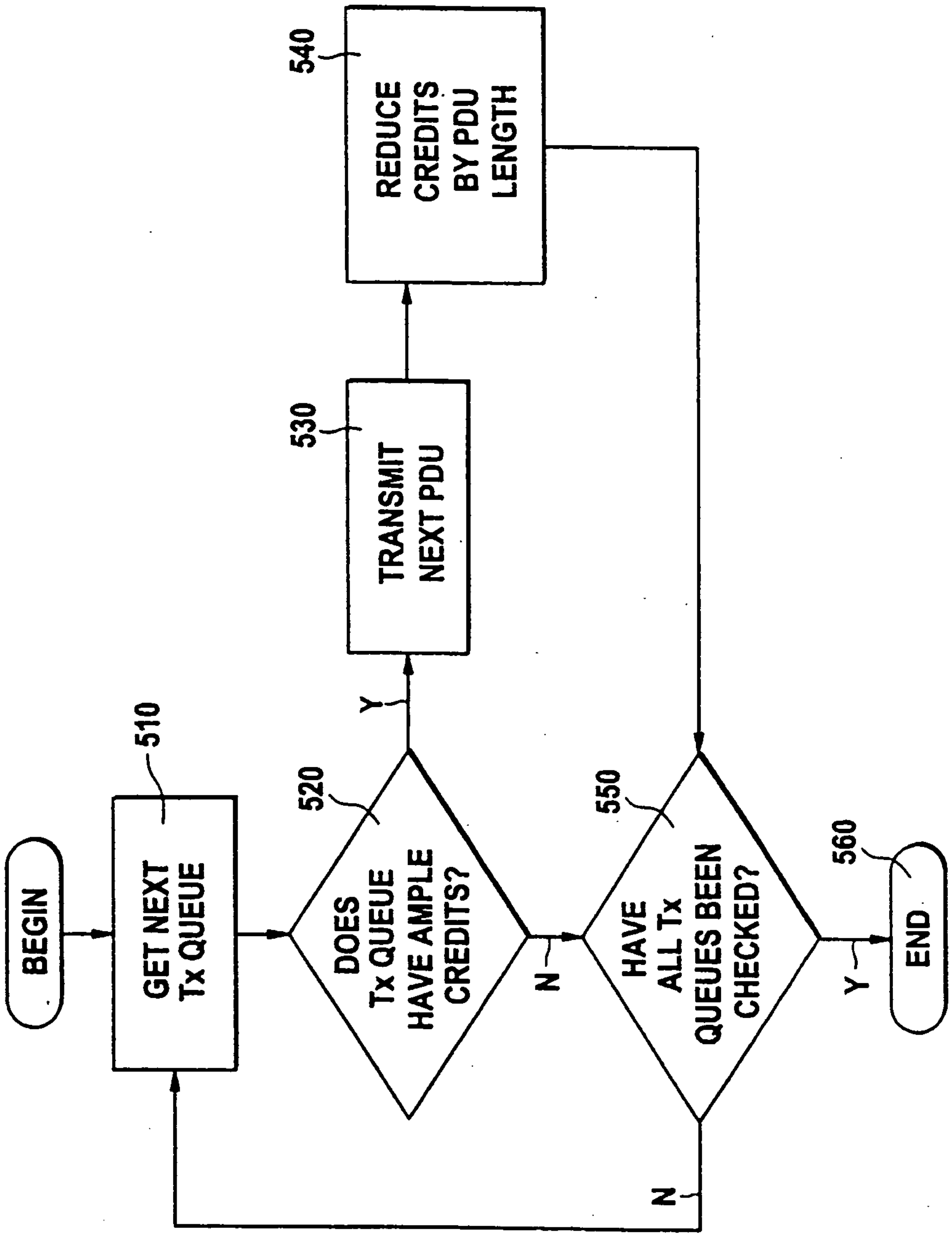

BEGIN
GET NEXT Tx QUEUE
510
DOES Tx QUEUE HAVE AMPLE CREDITS?
520
Y
N
TRANSMIT NEXT PDU
530
REDUCE CREDITS BY PDU LENGTH
540
HAVE ALL Tx QUEUES BEEN CHECKED?
550
Y
N
END
560

Fig. 5

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 99/04376

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 H04L12/56 H04Q11/04

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 H04L H04Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LYNN M A ET AL: "THE PRIORITY TOKEN BANK IN A NETWORK OF QUEUES" 1997 IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS, MONTREAL, JUNE 8 - 12, 1997, vol. 3, 8 June 1997 (1997-06-08), pages 1387-1391, XP000748872 INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERSISBN: 0-7803-3926-6 page 1387, right-hand column, line 45 - page 1388, left-hand column, line 9 page 1388, right-hand column, line 33 - line 37 page 1389, right-hand column, paragraph 4.3 ---	1,3-6,8, 12, 14-17, 19,22

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
1 September 1999	10/09/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Veen, G

1

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 99/04376

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	WO 97 14240 A (AISSAOUI MUSTAPHA ;LIAO RAYMOND RUI FENG (CA); NEWBRIDGE NETWORKS) 17 April 1997 (1997-04-17) claim 1 ---	1,12,22
A	TSANG D H K ET AL: "CREDIT-BASED FAIR QUEUEING FOR ATM NETWORKS" ELECTRONICS LETTERS, vol. 32, no. 25, 5 December 1996 (1996-12-05), page 2306/2307 XP000685308 ISSN: 0013-5194 page 2306, right-hand column -----	1,12,22

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/EP 99/04376

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9714240 A	17-04-1997	AU 7123596 A	30-04-1997
		CA 2234621 A	17-04-1997
		EP 0872088 A	21-10-1998